

# Machine Learning Group:

- Trust and Accuracy of Data
- Data Validation
- Remove roadblocks for progress
- Came up with 17 items
- Having access to data
- Need an association like the ASHA to publish information around these areas with goal of becoming more transparent
- Collaboration between computer scientists and survey researchers is needed

- 1) Cost Savings from Big Data??
- 2) Trust & Accuracy from Big Data?
- 3) Best Practice Methods
- 4) Transparency of Measurement,
- 5) Preprocessing (Cleaning & Imputation)
- 6) Transparency of Imputation
- 7) Multiple Use-cases (standard of "good")
- 8) Non-response Bias Correction
- 9) Software manipulation of Data
- 10) Process of reporting Small-end data
- 11) Data Quality issues in Training/Testing/  
Causal Inference

12) Multiple datasources to overcome problems from Big Data

13) Benchmarks or statistical processes to validate results

14) Engagement b/w ML folks and the problem being solved

15) Fitness between data being used and intended outcome

16) Privacy ~~collection~~

17) Sentiment  $\Rightarrow$  Outcome